

1 **CLAIMS**

2

3 1. A method comprising:

4 selecting, for each of a plurality of objects stored on a plurality of

5 computers in a network, a portion of object information corresponding to the

6 object;

7 using a stochastic partitioning process to identify which of the plurality of

8 computers to communicate the object information to for identification of

9 potentially identical objects on the plurality of computers.

10

11 2. A method as recited in claim 1, wherein the object information

12 comprises file information and wherein each of the plurality of objects comprises a

13 file in a file system.

14

15 3. A method as recited in claim 1, wherein the stochastic partitioning

16 process comprises a fully distributed stochastic partitioning process including:

17 comparing, for each of the plurality of computers, the selected portion to a

18 portion of a computer identifier associated with the computer;

19 identifying which of the computer identifiers have portions matching the

20 selected portion of the object information; and

21 communicating the object information to each of the computers associated

22 with a computer identifier having a portion matching the selected portion of the

23 object information.

24

25

1           4.    A method as recited in claim 1, wherein the stochastic partitioning  
2 process comprises a group-based system using directory services process  
3 including:

4           accessing an object information portion to computer mapping on a remote  
5 computer; and

6           identifying one or more computers to receive the object information based  
7 at least in part on the accessed mapping.

8  
9           5.    A method as recited in claim 1, wherein the stochastic partitioning  
10 process comprises a multi-level stochastic partitioning process including:

11           grouping, into a plurality of groups, selected ones of the plurality of  
12 computers, wherein the grouping is based at least in part on the number of the  
13 plurality of computers in the network that the computer using the stochastic  
14 partitioning process is aware of; and

15           identifying which of the selected ones of the plurality of computers to  
16 communicate the object information to, wherein the identifying is based at least in  
17 part on comparing the selected portion of the object information to a portion of a  
18 computer identifier of one or more of the selected ones of the plurality of  
19 computers.  
20  
21  
22  
23  
24  
25

1       6. One or more computer-readable media having stored thereon a  
2 plurality of instructions that, when executed by one or more processors of a  
3 computer that is part of a plurality of computers in a network, causes the one or  
4 more processors to perform the following acts:

5       selecting a portion of file information corresponding to a file stored on one  
6 of the plurality of computers;

7       comparing, for each of the plurality of computers, the selected portion to a  
8 portion of a computer identifier associated with the computer;

9       identifying which of the computer identifiers have portions matching the  
10 selected portion of the file information; and

11       communicating, for identification of potentially identical files stored on the  
12 plurality of computers, the file information to each of the computers associated  
13 with a computer identifier having a portion matching the selected portion of the  
14 file information.

15  
16       7. One or more computer-readable media as recited in claim 6, wherein  
17 the selected portion of the file information comprises a set of least significant bits  
18 of the file information.

19  
20       8. One or more computer-readable media as recited in claim 6, wherein  
21 for each of the plurality of computers the portion of the computer identifier  
22 associated with the computer is a set of least significant bits of the computer  
23 identifier.

1           **9.**     One or more computer-readable media as recited in claim 6, wherein  
2 the comparing includes comparing the selected portion to a portion of a computer  
3 identifier associated with the one computer.

4  
5           **10.**    One or more computer-readable media as recited in claim 6, wherein  
6 the selected portion of the file information includes a plurality of bits of the file  
7 information, wherein each portion of a computer identifier includes a plurality of  
8 bits of the computer identifier, and wherein the locations of the plurality of bits in  
9 the file information correspond to the locations of the plurality of bits of the  
10 computer identifiers.

11  
12           **11.**    One or more computer-readable media as recited in claim 6, further  
13 comprising:

14           receiving, from one of the plurality of computers, other file information;  
15           comparing the received file information to a file information database  
16 stored at the computer; and

17           determining that a file corresponding to the received file information is  
18 potentially identical to another file on another of the plurality of computers if the  
19 received file information matches file information in the database.

20  
21           **12.**    One or more computer-readable media as recited in claim 6, wherein  
22 the file information is a semi-unique value based at least in part on the data in the  
23 file.

1       **13.**   One or more computer-readable media as recited in claim 12,  
2 wherein the file information includes a hash value generated by applying a hash  
3 function to the file.

4  
5       **14.**   One or more computer-readable media as recited in claim 6, wherein  
6 the file information is based at least in part on one or more characteristics of the  
7 file.

8  
9       **15.**   One or more computer-readable media as recited in claim 14,  
10 wherein the characteristics include a file size and a file type.

11  
12       **16.**   One or more computer-readable media as recited in claim 6, wherein  
13 the communicating further comprises communicating a file identifier  
14 corresponding to the file along with the file information.

15  
16       **17.**   One or more computer-readable media as recited in claim 6, wherein  
17 the size of the portion of the file information is based at least in part on a count of  
18 computers that the one computer is aware of in the network.

19  
20       **18.**   One or more computer-readable media as recited in claim 6, wherein  
21 the size of the portion of the file information is based at least in part on an average  
22 number of computers that a particular file identifier should be communicated to.

1           **19.**   One or more computer-readable media as recited in claim 6, further  
2 comprising:

3           determining a size of the portion of the file information;

4           wherein the size of the portion of the file information need not be the same  
5 as a size determined by others of the plurality of computers.

6  
7           **20.**   One or more computer-readable media as recited in claim 6, wherein  
8 a value  $W$  represents the size of the portion of the file information, wherein a value  
9  $M$  represents a count of computers that the one computer is aware of in the  
10 network, wherein a value  $R$  is a system configuration value calculated based on an  
11 average number of computers that a particular file identifier should be  
12 communicated to, wherein  $\lg$  is a base 2 logarithm function, wherein floor  
13 brackets indicate the largest integer that is no greater than the enclosed value, and  
14 wherein the value  $W$  is determined as follows:

15           
$$W = \left\lfloor \lg \frac{M}{R} \right\rfloor.$$

16  
17           **21.**   One or more computer-readable media as recited in claim 6, further  
18 comprising repeating the selecting, comparing, identifying, and communicating  
19 for every file stored at the one computer.

20  
21           **22.**   One or more computer-readable media as recited in claim 6, further  
22 comprising repeating the selecting, comparing, identifying, and communicating  
23 for every file stored at the one computer that exceeds a threshold size.  
24  
25

1           **23.**    A method comprising:

2           generating an imprint for an object stored at a computer, wherein the  
3 imprint comprises a first set of bits of object information corresponding to the  
4 object;

5           identifying one or more additional computers each having a computer  
6 identifier that includes a second set of bits that match the imprint; and

7           sending the object information to each of the identified one or more  
8 additional computers.

9  
10          **24.**    A method as recited in claim 23, wherein the object comprises a file  
11 and the object information comprises file information.

12  
13          **25.**    A method as recited in claim 23, wherein the first set of bits  
14 comprises the least significant bits of the object information, and wherein the  
15 second set of bits comprises the least significant bits of the computer identifier.

16  
17          **26.**    A method as recited in claim 23, wherein the object information is a  
18 semi-unique value based at least in part on the data in the object.

19  
20          **27.**    A method as recited in claim 23, wherein the communicating further  
21 comprises communicating an object identifier corresponding to the object along  
22 with the object information.

1           **28.**   A method as recited in claim 23, further comprising storing the  
2 object information in an object information database stored at the computer.

3  
4           **29.**   A method as recited in claim 23, further comprising:  
5       receiving object information corresponding to an object stored on another  
6 computer;  
7       comparing the received object information to an object information  
8 database;  
9       checking whether the received object information matches any of the object  
10 information in the database; and  
11       determining that two potentially identical objects exist if the received object  
12 information matches any of the object information in the database.

13  
14           **30.**   A method as recited in claim 23, wherein the locations of the first  
15 set of bits correspond to the locations of the second set of bits.

16  
17           **31.**   A method as recited in claim 23, wherein the number of bits in the  
18 first set of bits is based at least in part on a count of computers that the computer is  
19 aware of in the network.

20  
21           **32.**   A method as recited in claim 23, wherein the number of bits in the  
22 first set of bits is based at least in part on an average number of computers that a  
23 particular object identifier should be communicated to.



1           **33.**    A system comprising:  
2           an interface configured to allow the system to communicate with a plurality  
3 of other computers; and  
4           a forwarding location determination module, coupled to the interface,  
5 configured to identify one or more of the plurality of computers to communicate  
6 file information corresponding to a file to, for identification of potentially identical  
7 files in the system, by,  
8                 identifying a set of bits of file information associated with a file, and  
9                 identifying ones of the one or more computers that each have a  
10 computer identifier that has a set of bits that match the set of bits of the file  
11 information.

12  
13           **34.**    A system as recited in claim 33, wherein the number of bits in the  
14 set of bits is based at least in part on a count of computers coupled to the system  
15 that the system is aware of.

16  
17           **35.**    A system as recited in claim 33, wherein the number of bits in the  
18 set of bits is based at least in part on an average number of computers that a  
19 particular file identifier should be communicated to.

20  
21           **36.**    A system as recited in claim 33, further comprising a file  
22 information comparison module coupled to a file information database, wherein  
23 the file information comparison module is configured to,  
24                 receive, from one of the plurality of computers, other file information,  
25

1 compare the received file information to file information in the database,  
2 and  
3 determine that a file corresponding to the received file information is  
4 potentially identical to another file on another of the plurality of computers if the  
5 received file information matches file information in the database.  
6  
7

8 **37.** A method comprising:  
9 identifying an imprint for an object stored at a computer, wherein the  
10 imprint is a set of bits of object information corresponding to the object;  
11 accessing an imprint to computer mapping;  
12 identifying one or more computers to receive the object information based  
13 at least in part on the accessed mapping; and  
14 sending the object information to at least one of the identified one or more  
15 computers.  
16

17 **38.** A method as recited in claim 37, wherein the object comprises a file  
18 and wherein the object information comprises file information.  
19

20 **39.** A method as recited in claim 37, wherein the one or more computers  
21 includes the computer at which the object is stored.  
22

23 **40.** A method as recited in claim 37, wherein the accessing comprises  
24 accessing a locally stored imprint to computer mapping.  
25

1           **41.**    A method as recited in claim 37, wherein the accessing comprises  
2 accessing an imprint to computer mapping stored at another computer.

3  
4           **42.**    A method as recited in claim 37, wherein the identifying one or  
5 more computers comprises identifying one computer to receive the object  
6 information based at least in part on the accessed mapping, and wherein the  
7 sending comprises sending the object information to the one identified computer  
8 so that the one identified computer can forward the object information to one or  
9 more other computers responsible for maintaining the object information.

10  
11           **43.**    A method as recited in claim 37, wherein the imprint to computer  
12 mapping comprises an imprint to computer identifier mapping.

13  
14           **44.**    A method as recited in claim 37, wherein the sending comprises  
15 sending the object information to each of the identified one or more computers.

16  
17           **45.**    A method as recited in claim 37, wherein the object information is a  
18 semi-unique value based at least in part on the data in the object.

19  
20           **46.**    A method as recited in claim 37, wherein the object information is  
21 based at least in part on one or more characteristics of the object.

1           47.    A method as recited in claim 37, wherein the size of the imprint is  
2 based at least in part on a count of other computers in a network coupled to the  
3 computer that the computer is aware of.

4  
5           48.    A method as recited in claim 37, wherein the size of the imprint is  
6 based at least in part on an average number of computers that a particular object  
7 identifier should be communicated to.

8  
9           49.    A method as recited in claim 37, further comprising:  
10 receiving object information corresponding to another object;  
11 comparing the received object information to an object information  
12 database;  
13 checking whether the received object information matches any of the object  
14 information in the database; and  
15 determining that two potentially identical objects exist if the received object  
16 information matches any of the object information in the database.

17  
18           50.    One or more computer-readable media having stored thereon a  
19 plurality of instructions that, when executed by one or more processors of one of a  
20 plurality of computers in a network, causes the one or more processors to perform  
21 the following acts:

22           selecting a portion of file information corresponding to a file;  
23 identifying a mapping of the portion to one or more computers; and  
24  
25

1 communicating the file information to each of the identified one or more  
2 computers for identification of potentially identical files on the one or more  
3 computers.  
4

5 **51.** One or more computer-readable media as recited in claim 50,  
6 wherein the file information is a semi-unique value based at least in part on the  
7 data in the file.  
8

9 **52.** One or more computer-readable media as recited in claim 50,  
10 wherein the file information is based at least in part on one or more characteristics  
11 of the file.  
12

13 **53.** One or more computer-readable media as recited in claim 50,  
14 wherein the size of the portion of the file information is based at least in part on a  
15 count of the plurality of computers in the network that the one computer is aware  
16 of.  
17

18 **54.** One or more computer-readable media as recited in claim 50,  
19 wherein the size of the portion of the file information is based at least in part on an  
20 average number of computers in the network that a particular file identifier should  
21 be communicated to.  
22  
23  
24  
25

1       **55.** One or more computer-readable media as recited in claim 50,  
2 wherein the identifying comprises identifying the mapping by accessing a locally  
3 stored imprint to computer mapping.  
4

5       **56.** One or more computer-readable media as recited in claim 50,  
6 wherein the identifying comprises identifying the mapping by accessing an imprint  
7 to computer mapping stored at another computer.  
8

9       **57.** A system comprising:  
10 an interface configured to allow the system to communicate with a plurality  
11 of other computers; and  
12 a forwarding location determination module, coupled to the interface,  
13 configured to identify one or more of the plurality of other computers to  
14 communicate file information corresponding to a file to for identification of  
15 potentially identical files stored on the plurality of other computers by accessing a  
16 mapping of a portion of the file information to one or more computers.  
17

18       **58.** A system as recited in claim 57, wherein the file information is a  
19 semi-unique value based at least in part on the data in the file.  
20

21       **59.** A system as recited in claim 57, wherein the size of the portion of  
22 the file information is based at least in part on a count of computers coupled to the  
23 system that the system is aware of.  
24  
25

1           **60.**    A system as recited in claim 57, further comprising a local imprint  
2 to computer mapping, and wherein the forwarding location determination module  
3 is configured to identify the one or more computers to communicate the file  
4 information to by accessing the local imprint to computer mapping.

5  
6           **61.**    A system as recited in claim 57, wherein the forwarding location  
7 determination module is configured to identify the one or more computers to  
8 communicate the file information to by accessing an imprint to computer mapping  
9 stored at another computer.

10  
11          **62.**    A method comprising:  
12           receiving file information corresponding to a file stored at a computer;  
13           comparing the received file information to a file information database;  
14           checking whether the received file information matches any of the file  
15 information in the database;  
16           determining that two potentially identical files exist if the received file  
17 information matches any of the file information in the database; and  
18           forwarding the received file information to another computer for storage in  
19 a file information database at the other computer.

20  
21          **63.**    A method as recited in claim 62, wherein the file information is a  
22 semi-unique value based at least in part on the data in the file.  
23  
24  
25

1           64. One or more computer-readable media having stored thereon a  
2 plurality of instructions that, when executed by one or more processors, causes the  
3 one or more processors to perform the following acts:

4           receiving, from a requesting computer, a request for an imprint to computer  
5 mapping, wherein the imprint comprises a portion of object information  
6 corresponding to an object;

7           accessing a mapping database to identify one or more computers associated  
8 with the imprint; and

9           returning an identification of at least one of the one or more computers to  
10 the requesting computer.

11  
12           65. One or more computer-readable media as recited in claim 64,  
13 wherein the object comprises a file and wherein the object information comprises  
14 file information.

15  
16           66. One or more computer-readable media as recited in claim 64,  
17 wherein the at least one of the one or more computers includes the requesting  
18 computer.

19  
20           67. One or more computer-readable media as recited in claim 64,  
21 wherein the returning comprises returning an identification of all of the one or  
22 more computers to the requesting computer.



1           **68.**   One or more computer-readable media as recited in claim 64,  
2 wherein the file information is a semi-unique value based at least in part on the  
3 data in the object.  
4

5           **69.**   One or more computer-readable media as recited in claim 64,  
6 wherein the size of the portion of the object information is based at least in part on  
7 a count of computers coupled to the requesting computer that the requesting  
8 computer is aware of.  
9

10          **70.**   A method, implemented in a computer that is one of a plurality of  
11 computers in a network, the method comprising:

12           grouping, into a plurality of groups, selected ones of the plurality of  
13 computers, wherein the grouping is based at least in part on the number of the  
14 plurality of computers in the network that the computer is aware of;

15           selecting a portion of object information corresponding to an object ; and  
16           identifying which of the selected ones of the plurality of computers to  
17 communicate the object information to for identification of potentially identical  
18 objects stored on the plurality of computers, wherein the identifying is based at  
19 least in part on comparing the selected portion of the object information to a  
20 portion of a computer identifier of one or more of the selected ones of the plurality  
21 of computers.  
22

23          **71.**   A method as recited in claim 70, wherein the object is stored at the  
24 computer.  
25

1           72.    A method as recited in claim 70, wherein the object is stored at  
2 another of the plurality of computers.

3  
4           73.    A method as recited in claim 70, wherein the plurality of groups  
5 comprise two groups.

6  
7           74.    A method as recited in claim 70, further comprising:  
8 receiving object information corresponding to another object;  
9 comparing the received object information to an object information  
10 database;  
11 checking whether the received object information matches any of the object  
12 information in the database; and  
13 determining that two potentially identical objects exist if the received object  
14 information matches any of the object information in the database.

15  
16           75.    A method as recited in claim 70, wherein a size of the portion of the  
17 object information is further based at least in part on an average number of  
18 computers that a particular object identifier should be communicated to.

19  
20           76.    A method as recited in claim 70, wherein the object information is a  
21 semi-unique value based at least in part on the data in the object.

1           77.     A method as recited in claim 70, wherein the size of the portion of  
2 the object information is the same as the size of the portions of the computer  
3 identifiers.

4  
5           78.     A method as recited in claim 70, wherein the portion of object  
6 information comprises a set of least significant bits of the object information.

7  
8           79.     A method as recited in claim 78, wherein the number of bits in the  
9 set of least significant bits is based at least in part on the number of other  
10 computers in the network that the computer is aware of.

11  
12           80.     A method as recited in claim 70, wherein the grouping comprises  
13 identifying as a first group one or more of the plurality of computers having  
14 computer identifiers each with a portion that is the same as a portion of a computer  
15 identifier of the computer.

16  
17           81.     A method as recited in claim 80, wherein the identifying comprises:  
18           checking whether all of the bits in the portion of the computer identifier  
19 match all of the bits in the selected portion of the object information; and  
20           identifying, as computers to communicate the object information to, the  
21 computers in the first group if all of the bits in the portion of the computer  
22 identifier match all of the bits in the selected portion of the object information.

1           **82.**     A method as recited in claim 70, wherein the grouping comprises:  
2           identifying as a first group one or more of the plurality of computers having  
3 computer identifiers each with a portion that is the same as a portion of a computer  
4 identifier of the computer; and  
5           identifying as a second group one or more of the plurality of computers that  
6 do not have computer identifiers each with the portion that is the same as the  
7 portion of the computer identifier of the computer, but that do have computer  
8 identifiers each with a first subset of the portion that is the same as a first subset of  
9 the portion of the computer identifier of the computer.

10  
11           **83.**     A method as recited in claim 82, wherein the first subset of each  
12 portion comprises the even bits of the portion.

13  
14           **84.**     A method as recited in claim 70, wherein the grouping comprises:  
15           identifying as a first group one or more of the plurality of computers having  
16 computer identifiers each with a portion that is the same as a portion of a computer  
17 identifier of the computer;  
18           identifying as a second group one or more of the plurality of computers that  
19 do not have computer identifiers each with the portion that is the same as the  
20 portion of the computer identifier of the computer, but that do have computer  
21 identifiers each with a first subset of the portion that is the same as a first subset of  
22 the portion of the computer identifier of the computer; and  
23           identifying as a third group one or more of the plurality of computers that  
24 do not have computer identifiers each with the first subset of the portion that is the  
25 same as the first subset of the portion of the computer identifier of the computer,

1 but that do have computer identifiers each with a second subset of the portion that  
2 is the same as a second subset of the portion of the computer identifier of the  
3 computer.

4  
5 **85.** A method as recited in claim 84, wherein the first subset of each  
6 portion comprises the even bits of the portion and the second subset of each  
7 portion comprises the odd bits of the portion.

8  
9 **86.** A method as recited in claim 70, wherein the grouping comprises:  
10 identifying as one group one or more of the plurality of computers that have  
11 computer identifiers each with a subset of the portion of the computer identifier  
12 that is the same as a subset of the portion of the computer identifier of the  
13 computer.

14  
15 **87.** A method as recited in claim 86, wherein the identifying comprises:  
16 checking whether each bit in another subset of the portion of the computer  
17 identifier matches a corresponding bit in the selected portion of the object  
18 information; and

19 if each bit in the other subset of the portion of the computer identifier  
20 matches the corresponding bit in the selected portion of the object information,  
21 then identifying, as computers to communicate the object information to, the  
22 computers in the one group having a portion of their computer identifiers matching  
23 the corresponding bits in the selected portion of the object information.

1       **88.**     A method as recited in claim 86, wherein the identifying comprises:  
2       checking whether each bit in another subset of the portion of the computer  
3 identifier matches a corresponding bit in the selected portion of the object  
4 information; and  
5       if one or more bits in the other subset of the portion of the computer  
6 identifier do not match the corresponding bit in the selected portion of the object  
7 information, then identifying, as computers to communicate the object information  
8 to, the computers in another group having the bits of a subset of their computer  
9 identifiers matching the selected portion of the object information.

10  
11       **89.**     A method as recited in claim 70, wherein the grouping comprises:  
12       identifying as a first group one or more of the plurality of computers having  
13 computer identifiers each with a portion that is the same as a portion of a computer  
14 identifier of the computer;  
15       identifying as a second group one or more of the plurality of computers that  
16 do not have computer identifiers each with the portion that is the same as the  
17 portion of the computer identifier of the computer, but that do have computer  
18 identifiers each with a first subset of the portion that is the same as a first subset of  
19 the portion of the computer identifier of the computer;  
20       identifying as a third group one or more of the plurality of computers that  
21 do not have computer identifiers each with the portion that is the same as the  
22 portion of the computer identifier of the computer, but that do have computer  
23 identifiers each with a second subset of the portion that is the same as a second  
24 subset of the portion of the computer identifier of the computer;  
25

1 checking whether all of the bits in the portion of the computer identifier  
2 match all of the bits in the selected portion of the object information;

3 if all of the bits in the portion of the computer identifier match all of the bits  
4 in the selected portion of the object information, then identifying, as computers to  
5 communicate the object information to, the computers in the first group; and

6 if all of the bits in the portion of the computer identifier do not match all of  
7 the bits in the selected portion of the object information, then,

8 checking whether the bits in the second subset of the portion of the  
9 object information match the bits in the second subset of the computer  
10 identifier,

11 if the bits in the second subset of the portion of the object  
12 information match the bits in the second subset of the computer identifier,  
13 then identifying, as computers to communicate the object information to,  
14 the computers in the third group having their computer identifiers matching  
15 the selected portion of the object information, and

16 if the bits in the second subset of the portion of the object  
17 information do not match the bits in the second subset of the computer  
18 identifier, then identifying, as computers to communicate the object  
19 information to, the computers in the second group having the subset of bits  
20 in their computer identifiers matching the corresponding bits in the selected  
21 portion of the object information.

1           **90.**   One or more computer-readable media having stored thereon a  
2 plurality of instructions that, when executed by one or more processors of a  
3 computer that is one of a plurality of computers in a network, causes the one or  
4 more processors to perform the following acts:

5           selecting ones of the plurality of computers to populate a plurality of  
6 groups, wherein the selecting is based at least in part on the number of computers  
7 in the network that the computer is aware of;

8           selecting a plurality of bits of file information corresponding to a file; and  
9           identifying which of the selected ones of the plurality of computers to  
10 communicate the file information to for identification of potentially identical files  
11 on the plurality of computers, wherein the identifying is based at least in part on  
12 comparing the selected plurality of bits of the file information to a corresponding  
13 plurality of bits of a computer identifier of one or more of the selected ones of the  
14 plurality of computers.

15  
16           **91.**   One or more computer-readable media as recited in claim 90,  
17 wherein the file is stored at the computer.

18  
19           **92.**   One or more computer-readable media as recited in claim 90,  
20 wherein the file is stored at another of the plurality of computers.

21  
22           **93.**   One or more computer-readable media as recited in claim 90,  
23 wherein the plurality of instructions further cause the one or more processors to  
24 perform the following acts:

25           receiving file information corresponding to another file;



1 comparing the received file information to a file information database;  
2 checking whether the received file information matches any of the file  
3 information in the database; and  
4 determining that two potentially identical files exist if the received file  
5 information matches any of the file information in the database.  
6

7 **94.** One or more computer-readable media as recited in claim 90,  
8 wherein the selecting comprises identifying as a first group one or more of the  
9 plurality of computers having computer identifiers each with a plurality of bits that  
10 is the same as a plurality of bits of a computer identifier of the computer.  
11

12 **95.** One or more computer-readable media as recited in claim 94,  
13 wherein the identifying comprises:

14 checking whether all of the plurality of bits of the computer identifier  
15 match all of the plurality of bits of the file information; and

16 identifying, as computers to communicate the file information to, the  
17 computers in the first group if all of the plurality of bits of the computer identifier  
18 match all of the plurality of bits of the file information.  
19

20 **96.** A method as recited in claim 90, wherein the selecting comprises:

21 identifying as a first group one or more of the plurality of computers having  
22 computer identifiers each with a plurality of bits that is the same as a plurality of  
23 bits of a computer identifier of the computer; and

24 identifying as a second group one or more of the plurality of computers that  
25 do not have computer identifiers each with a plurality of bits that is the same as the

1 plurality of bits of the computer identifier of the computer, but that do have  
2 computer identifiers each with a first subset of the plurality of bits that is the same  
3 as a first subset of the plurality of bits of the computer identifier of the computer.  
4

5 97. A system, coupled to a plurality of computers, the system  
6 comprising:

7 an interface configured to allow the system to communicate with the  
8 plurality of computers; and

9 a forwarding location determination module, coupled to the interface,  
10 configured to identify one or more of the plurality of computers to communicate  
11 the file information for a file to, for identification of potentially identical files  
12 stored on the plurality of computers by,

13 grouping, into a plurality of groups, selected ones of the plurality of  
14 computers, wherein the grouping is based at least in part on the number of  
15 the plurality of computers in the network that the system is aware of,

16 selecting a portion of the file information, and

17 comparing the selected portion of the file information to a portion of  
18 a computer identifier of one or more of the selected ones of the plurality of  
19 computers.  
20  
21  
22  
23  
24  
25